Understanding the composite dimensions of the EQ-5D: an experimental approach

Rebecca McDonald[1] (University of Birmingham), Timothy L. Mullett (University of Warwick), Aki Tsuchiya (University of Sheffield)

Abstract

The EQ-5D(-5L) includes two composite dimensions: "Pain or Discomfort" (P/D) and "Anxiety or Depression" (A/D), which involves an inherent ambiguity. Little is known about how these composite dimensions are interpreted across contexts where (i) individuals self-report their own health; and (ii) individuals value stylised health states. We detail the nature of the ambiguity and present experimental evidence from two large online surveys (n=1007 and n=1415). In one survey, individuals reported both their current health and their health at the time they felt the worst because of their health. In the other, they valued stylised EQ-5D states using Discrete Choice Experiments with duration as an attribute. In both surveys, participants were randomised into treatments in which the presentation of one of the composite dimensions was altered, or a control. Our results suggest (1) In self-report, use of the composite dimensions differs across the dimensions, with P/D used mainly to report Pain, but A/D used mainly to mean the more severe component of Anxiety and Depression. (2) In valuation, Pain was perceived to be worse than Discomfort at the same level, and Depression was perceived to be worse than Anxiety at the same level. (3) In valuation, the composite dimension P/D was interpreted to mean Pain, whilst the composite dimension A/D was interpreted to lie between Anxiety and Depression. We conclude that care must be taken when interpreting responses to existing health (or wellbeing) descriptive systems that rely on composite dimensions, and that caution should be applied when designing new ones.

Keywords: EQ-5D, self-reported health, health state valuation, composite dimensions

Declarations of interest: none.

[1] Correspondence to r.l.mcdonald@bham.ac.uk Department of Economics, University House, Edgbaston Campus, University of Birmingham, UK, B152TT

Introduction

The EQ-5D instrument (Brooks, 1996; Herdman et al, 2011) is widely used in the classification and valuation of different health states. These health states and their values underpin priority setting in health care, with real impacts on health across the population. The appropriateness of health care resource allocation therefore relies in part on the accuracy with which the EQ-5D captures the preferences of members of the public over relevant aspects of health. The core assumption is that the underlying perceptions of different health states are consistently and reliably measured and communicated by the EQ-5D classification system. To judge this assumption, researchers must develop a detailed understanding about how the EQ-5Dworks, both as a tool for self-reporting experienced health states and as a system for describing health states to the general population in valuation exercises.

This paper highlights a fundamental ambiguity in the EQ-5D system: it is not possible to logically determine how the levels of a composite dimension, viz. “Pain or discomfort” and “Anxiety or Depression”, ought to be interpreted. We clarify the nature of this ambiguity, and explore how the composite dimensions are actually used in self-reporting own health and interpreted in valuation exercises.

The ambiguity arises because a composite dimension essentially combines two different aspects of a health state into a single dimension. In the context of self-reporting own health using EQ-5D, someone with moderate anxiety and no depression could be expected to self-report “moderate problems with Anxiety or Depression”, but so could a second person with moderate anxiety and moderate depression, as could a third with no anxiety and moderate depression. The EQ-5D instrument is therefore fundamentally incapable of distinguishing between the health of these three people.

More generally, even if self-reporting behaviour always used the composite to report the level of the component with more severe problems, it is not possible to logically determine how the levels of a composite dimension ought to be interpreted. Moreover, little is known about how these composite

dimensions are actually used by members of the public when self-reporting their experienced health.

Furthermore, in the context of health state valuation where individuals value stylised health states described using EQ-5D, it is not clear how they interpret different levels of the composite dimensions. Someone presented with a health state including moderate Anxiety or Depression may interpret this as moderate anxiety and no depression, or moderate depression and no anxiety, or some other combination. If systematic differences exist in the way the composite dimensions are used between these self-report and valuation contexts, then health state values used in economic evaluations would be systematically biased. For example, it is conceivable that individuals with moderate anxiety and no depression self-report "moderate Anxiety or Depression" while individuals valuing a stylised health state with "moderate Anxiety or Depression" interpret this as moderate depression and no anxiety. If so, and if moderate depression is considered to be worse than moderate anxiety, there will be systematic overvaluation of health states involving moderate Anxiety or Depression. Further discussion of the possible interpretations and use of the EQ-5D composite dimensions is provided below. The key point is that there are multiple logically consistent but mutually incompatible interpretations of a given severity level of a composite dimension.

This paper examines how the composite dimensions are used in the contexts of self-reporting own health and of valuation exercises. We take an experimental approach, varying the presentation of the composite dimensions between subjects. To explore participants' use of the dimensions, treatments were designed in which either the Pain or Discomfort (or P/D for short) dimension or the Anxiety or Depression (A/D) dimension was altered. In some of the altered presentations, one of the composite dimension's components was not mentioned at all. In other presentations, the composite dimension was presented as two separate dimensions. This approach allows a wide range of possible interpretations of the composite dimensions to be investigated.

Literature

Previous evidence suggests that participants interpret the components of the composite dimensions to represent distinct concepts. For example, Bryan et al (2005) explored the interpretation of the A/D composite dimension. In a focus group study, A/D was presented as two separate dimensions in a three-level EQ-6D. Their qualitative results suggested that respondents tended to "interpret anxiety and depression as distinct and independent concepts". The authors also presented a quantitative study where patients self-reported their health using EQ-5D alongside other clinical measures. The correlation coefficients for the A/D item against clinical measures of depression were similar to the correlation coefficients for the A/D item against clinical measures of anxiety, and the authors interpreted this as evidence to support the use of the composite. However, they did not examine the effects of splitting A/D in the quantitative study. Furthermore, they considered only self-report data, and so any differences between self-report and valuation contexts were not accounted for.

Our approach includes what is essentially bolting-off components of the composite EQ-5D dimensions. Whilst such bolting off has until now only been considered by Tsuchiya et al (2019), a considerable literature exists that bolts *on* a dimension to the EQ-5D, including cognition (Krabbe et al, 1999; Wolfs et al, 2007), sleep (Yang et al, 2013), vision (Longworth et al, 2014), hearing (Longworth et al, 2014) and tiredness (Longworth et al, 2014). Studies repeatedly find that including a dimension with "no problems" may change the valuation of a health state (also see Brazier et al, 2011, which bolted on Pain or Discomfort to an asthma-specific preference-based instrument). This violates an implicit assumption of preference-based health state classification instruments, namely, that any unmentioned dimensions have no problems. Instead, explicitly stating that a dimension has no problems appears to generate different valuations compared to not mentioning the dimension. While our own approach is different, the conclusions of the bolt-on valuation literature might suggest that splitting a composite dimension into two when it had no problems might change the value of the health state compared to the unaltered version, both (a) if we keep both components; or (b) if we drop one or the other.

McDonald and Mullett (2020) investigated the effect of splitting P/D and A/D, whilst simultaneously collapsing Mobility and Usual Activities into a composite dimension. They found that splitting a dimension increased its importance in determining which health state was preferred in pairwise choice, and collapsing two dimensions into one reduced their importance. They concluded that individuals have a tendency towards equally weighting attributes in a multi-attribute choice. However, they were unable to examine the effect of dropping components and did not consider the self-report context.

Finally, Tsuchiya et al (2019) examined, amongst other things, the effect of splitting the composite dimensions of the EQ-5D and presenting both components separately in place of the composite dimension, so forming EQ-6D. Comparing the use of each level of each dimension in self-report, they showed that reports of "no problems" were more frequent when the composite dimensions were presented, compared to where the composites were split into two separate components. This implies individuals do not use the composite dimensions X/Y literally to mean "X or Y" in self-report. The effect of splitting the composites was more pronounced for A/D than for P/D, and the difference between the two composites may arise because of the differences in the way the components relate to one another. Although pain is commonly interpreted as a more severe form of discomfort (for example, in the well-established McGill Pain Questionnaires (Melzack, (1975; 1987)); and indicative evidence in Macran and Kind (2000)), there is evidence to suggest that anxiety and depression are entirely separate concepts. For a more detailed examination of the argument, see Bryan et al (2005).

Furthermore, Tsuchiya et al showed that in valuation tasks, the coefficients for the composite dimensions are related to, but not identical to, the sum of the coefficients on the components when both are presented separately. The patterns of their data suggest that splitting the composite has a different effect for P/D than for A/D, illustrating that we do not fully understand the way composite dimensions are used in the EQ-5D.

We present the first dedicated study to investigate the inherent ambiguity in EQ-5D regarding the P/D and the A/D composite dimensions across the full spectrum of possible interpretations by systematically varying the way the components are presented. Our specific aims were to ask:

(1) How are the P/D and A/D dimensions interpreted in self-reporting of own health?

(2) How are the P/D and A/D dimensions interpreted in valuation of stylised health states?

The results suggest that there are differences between the interpretations of the composite dimensions between P/D versus A/D. The interpretation also differs between self-report and valuation tasks for A/D, but our participants applied their interpretations of P/D consistently across the self-report and valuation tasks.

A theory on the use of composite dimensions

Table 1 sets out a stylised scheme that gives a series of logically possible interpretations of a self-reported level on a composite dimension under the assumption that the composite is used to report the severity level for the component on which the most severe problems are reported. It reports what the potential underlying levels of each component could be, for a given severity level of the composite. We only use the first, third and the fifth levels of EQ-5D-5L, since these are sufficient to illustrate our point. The only unambiguous composite dimension is given in row i: "No problems with X or Y", which must mean no problems on either component. However, if the respondent self-reports having "Moderate problems with X or Y", there are at least three potential combinations of the underlying components (rows ii - iv). Worse still, "Extreme problems" has five possible interpretations (rows v-ix). This demonstrates that even if self-reporting behaviour perfectly follows this pattern, it is not possible to logically determine how the levels of a composite dimension ought to be interpreted.

With all five levels of EQ-5D-5L, the total number of potential combinations per composite dimension expands from nine to 25 (one for level 1; three for level 2; five for level 3; seven for level 4; and nine for level 5). Since there are two composite dimensions, in effect, if a respondent

self-reports level 3 for both composite dimensions, this could logically mean any one of 25 (= 5 x 5) possible combinations of the four components.

Table 1 Ambiguity of interpretation of EQ5D showing multiple equally logical but mutually exclusive interpretations

Row	Report on Composite Dimension X or Y	Experience on component X	Experience on component Y
i	No problems	No problems	No problems
ii		No problems	Moderate problems
iii	Moderate problems	Moderate problems	No problems
iv		Moderate problems	Moderate problems
v		No problems	Extreme problems
vi		Moderate problems	Extreme problems
vii	Extreme problems	Extreme problems	No problems
viii		Extreme problems	Moderate problems
ix		Extreme problems	Extreme problems

In the context of health state valuation, respondents may interpret the levels of the composite dimensions in the health states to be valued as: the level of one or the other component which they think is more important; the level of both components; or any other combination. The interpretation is likely to vary across respondents, and may not be stable across the valuation exercise or across the two composite dimensions.

Methods

Design

The study used adapted versions of the EQ-5D-5L instrument to collect data on self-reported current health, self-reported health at the time that the respondent felt the worst because of their

health (hereafter "worst recalled health"), and valuation of stylised health states. These were conducted over two phases: Phase 1 collected the full valuation data and some limited self-reported data; Phase 2 collected more thorough self-reported data. The analyses reported in this paper are based on the self-reported data from the second phase and the valuation data from the first phase. For all analyses that could be applied in both datasets, results of the self-reported data from the first phase are consistent with those from the second phase, and are available upon request.

Self-reported worst recalled health was included because self-reported current health of the general public in EQ-5D-5L typically has around a third of the sample reporting full health, meaning little variation in the data (for example, Golicki and Niewada, 2017; Hinz et al., 2014; McCaffrey et al., 2016). The inclusion of worst recalled health is expected to result in wider variation across individuals, providing the basis for more informative analyses.

Table 2 The seven versions manipulating the presentation of the composite dimensions

Version	Dimensions
Standard	Mob, UA, SC, P/D, A/D
Drop Dis	Mob, UA, SC, Pai, A/D
Drop Pai	Mob, UA, SC, Dis, A/D
Split Pai Dis	Mob, UA, SC, Pai, Dis, A/D
Drop Dep	Mob, UA, SC, P/D, Anx
Drop Anx	Mob, UA, SC, P/D, Dep
Split Anx Dep	Mob, UA, SC, P/D, Anx, Dep

In both phases, participants were randomised into one of seven versions that differed according to how the composite dimensions were presented. In the standard version, participants saw the usual presentation of the EQ-5D with the composite dimensions unaltered. In the Drop Dis version, Discomfort was dropped and only Pain was presented. In Drop Pai, Pain was dropped and only

Discomfort was presented. In Split Pai Dis, Pain *and* Discomfort were presented as separate dimensions, creating a 6D version. Drop Dep, Drop Anx and Split Anx Dep were constructed similarly. Table 2 summarises these versions.

To clarify the explanation, we will refer to Drop Dis, Drop Pai, Drop Dep and Drop Anx as "partial drop" versions and the non-dropped dimensions will be referred to as Pain only, Discomfort only and so on. Split Pai Dis and Split Anx Dep will be referred to as "split" versions and the dimensions will be referred to as Anxiety separate, Depression separate and so on.

Self-reported own health

After reading an information sheet and giving informed consent, participants' first task was to self-report their current health using the dimensions for their version. The question asked:

> "Please indicate which statements best describe your health TODAY."

To answer, participants selected the relevant severity statement for each dimension. Next, respondents were asked about their worst recalled health, following the approach of Devlin et al (2017). Specifically, the question asked:

> "To help you start thinking about how you feel about different areas of health, we would like you to think about the time that you felt the worst because of your health. Indicate which statements best describe your health during the time that you felt worst because of your health."

Next, respondents assessed their own general health as: excellent, very good, good, fair, or poor. Finally, participants were asked to self-report their health on the missing dimensions. Specifically, participants who were in the Partial Drop treatments self-reported their current health on the dropped dimension and on the composite dimension; participants who saw the composite dimensions self-reported their health on each component separately; and participants who were in the Split treatments self-reported their current health on the composite dimension. The process was

repeated for worst recalled health, generating complete information about self-reported current and worst recalled health.

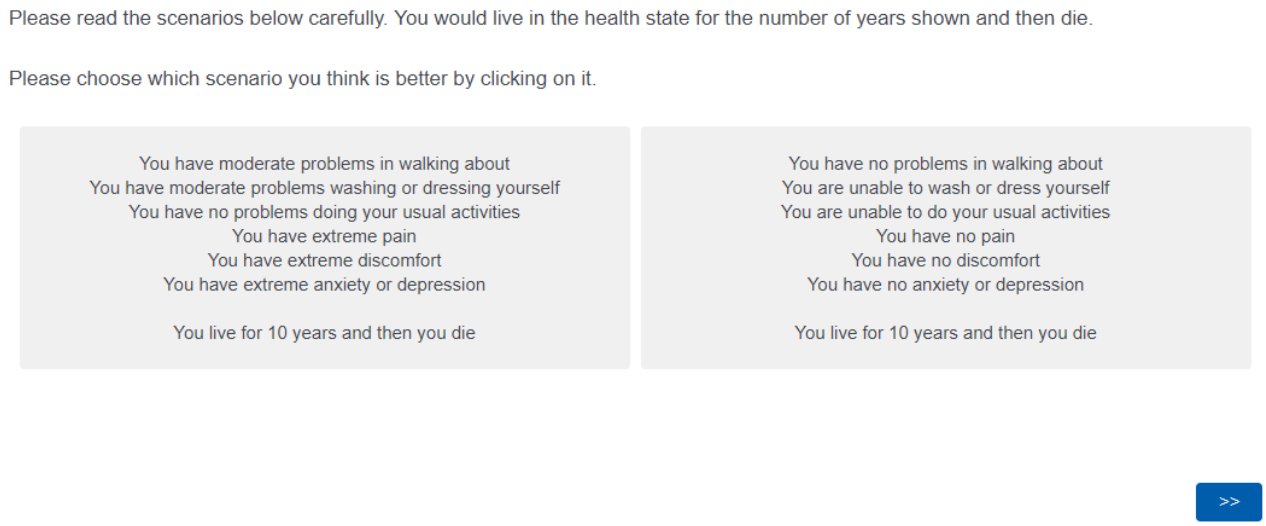
Health state valuation

A Discrete Choice Experiment with a duration attribute (DCE_{TTO}) was used for the health state valuations. This approach was developed by Bansback et al (2012) and refined in Bansback *et al.* (2014) and Mulhern *et al.* (2018). It combines the Discrete Choice Experiment with the Time Trade Off (TTO) to allow health state values to be elicited from paired choices, with values on a scale anchored at 1 for full health and 0 for a state equivalent to being dead (for a review, see Mulhern *et al.*, 2019). In our application, it involves participants making a series of pairwise choices between stylised health scenarios described by the five (or six) dimensions relevant to their version of the experiment, plus a duration dimension that could take the value 6 years, 8 years or 10 years, followed by death. Since the study had a methodological focus and did not aim to produce an alternative value set for EQ-5D-5L, for the DCE_{TTO} we restricted the levels of the health states to be level 1 (no problems), level 3 (moderate problems) or level 5 (extreme problems/unable), omitting "slight" and "severe" problems. This reduces the number of possible health state comparisons, and follows Tsuchiya et al (2019). An example task is provided in Figure 1 and the instructions for the DCE_{TTO} task are provided in Appendix 1.

Choice sets were selected using Ngene (Choice Metrics, 2012) with priors of zero and assuming that a conditional logit model was the true model. Ten balanced D-efficient designs were generated for 5D and 6D. Whilst the final designs cannot be directly compared between the 5D and 6D versions (they have different numbers of attributes), they were designed using the same process and were each maximally efficient. We selected 48 choice sets for the 5D versions (the standard and partial drop versions) and 60 for the 6D versions (the split versions). This allows us to estimate a model with linear duration, including main effects and interactions for each of the attribute levels and duration, which involves 21 and 25 parameters for the 5D and 6D cases, respectively. For each

participant in each version, 20 of these pairs were drawn and presented at random from the 48 or 60 possible tasks.

Figure 1 DCE_{TTO} task from Version 2c, splitting P/D into separate attributes



Before beginning the pairwise choices, participants faced four practice questions. Three involved dominance across relatively mild states: the first question held the severity levels constant but differed in duration; the second question had dominance on both duration and severity; and the third question held duration constant but one scenario was dominated on severity. In all three cases, if the dominated option was selected a warning pop-up appeared to explain the mistake. However, participants could continue with the dominated option selected, giving us a check on participants' understanding. The final practice question did not involve a dominated option.

After completing the main valuation tasks, respondents were invited to select all statements that applied to their experience of the health state valuation exercise.

Background characteristics

At the end, demographic information was collected, including gender, age, having experienced serious illness, employment status, education, and whether the participant was responsible for children under 18.

Recruitment

Participants for both phases were recruited through the Prolific.ac online participant pool. Members of this pool have varied ages, genders and incomes and are drawn from geographically diverse locations. We did not pre-screen on demographic characteristics or language, but restricted the sample to UK residents aged 18 or over. Importantly, the samples for both phases were drawn from the same population. In an information page, participants were informed that their data would be kept confidential and would not be linked to their identity. Ethical approval was granted by the University of Warwick's Humanities and Social Sciences Research Ethics Committee. Participants in phase 1 received £1.50 for their participation and participants in phase 2 received £0.80 (since their task was less time consuming). Phase 1 data were collected in July 2017 and phase 2 data were collected in August 2019.

Analytical approach

Self-report data

To analyse the data for self-reporting own health, we first conduct a test of 'literal' behaviour, establishing whether participants treated the composite dimensions as literally meaning "or" by comparing the incidence of "no problems" being reported, since these are the cases for which the composite and its components are unambiguously related (see Table 1 for details). In the event that these tests show that participants do not appear to treat the composite literally, we then conduct a test of self-reporting rules, which established what reporting rule best fits the data.

We hypothesise four potential self-reporting rules as follows:

1. The composite is used as the worse problem across the components (i.e. as an "or")

   e.g. with moderate problems on one component and extreme problems on the other, this rule would predict self-reporting extreme problems on the composite

2. The composite is used to self-report one of the two components:

   2A. the first mentioned component (i.e. Pain of P/D or Anxiety of A/D)

e.g. with moderate pain and extreme discomfort, this rule would predict self-reporting moderate problems on the composite

2B. the last mentioned component (i.e. Discomfort of P/D or Depression of A/D)

e.g. with extreme anxiety and moderate depression, this rule would predict self-reporting moderate anxiety or depression on the composite

3. The composite is used as an average (mean) across the components

e.g. with moderate problems on one component and extreme problems on the other, this rule would predict self-reporting severe problems on the composite

In the test of self-reporting rules, we set out exemplar response patterns for the four different reporting rules and we compare the degree to which the data differ from the exemplars. For each individual, based on their reported levels of Pain, Discomfort, Anxiety and Depression, we predict their P/D and A/D according to the four reporting rules. Then, for each individual and by each self-reporting rule, we take the difference between the actual and the predicted levels for P/D and A/D. Finally, for each composite dimension and by each of the self-reporting rules, we calculate the mean absolute difference between the prediction and the actual report, pooling across the individuals. This gives a quantified measure of the error in predictions. The self-reporting rule with the smallest error best represents participants' self-reporting behaviour.

To provide confidence intervals around the estimate of error for each self-reporting rule, we apply a bootstrapping approach. For each composite dimension, by each self-reporting rule, participants are randomly sampled with replacement and the mean absolute error calculated for 10,000 samples. The distribution of mean absolute errors is used to find the 95-percentile range.

DCE_{TTO} data

To analyse the valuation data from the DCE_{TTO}, we follow the approach taken by Bansback *et al.* (2012). The approach is to model participants' utility (μ_{ij}) where i denotes the participant and j denotes the health scenario being considered, such that j=1,2 are the scenarios considered in each

pairwise choice. Utility is modelled as a function of all possible attribute levels of the EQ-5D (or 6D in the split conditions), and duration. We let a vector of dummy variables for each possible attribute level be *x*, with "no problems" as the reference category. We model duration, *t*, as continuous. This gives the formula (drawn from Bansback *et al.* 2012):

$$\mu_{ij} = \alpha + \beta t_{ij} + \lambda' x_{ij} t_{ij} + \epsilon_{ij} \qquad (1)$$

In this model, α is a constant measuring the tendency to choose the specified option, *ceteris paribus*; β captures the preference for living in full health for one year, λ is the disutility associated with the levels of the attributes specified in the given health state when experienced for one year, and ϵ_{ij} is the error term, assumed to be iid. We make the standard assumption of constant proportional trade-offs in life years, modelling duration as a continuous, linear variable.

A mixed model logistic regression is estimated for each presentation version, to establish the factors influencing the choice of one health scenario over another. This involves specifying, for each comparison, the difference between the two options on the specified dimensions (duration and attribute levels) and modelling the probability that an option is chosen over another option given these differences. Dummy variables are used for the attribute severity levels (variable x in Eq. 1). If the health scenario presented on the left-hand side of the screen has the relevant severity level on a dimension, the value of the dummy is set to 1; if the scenario presented on the right has this severity level on this dimension, the dummy is set to -1. If the health scenarios have the same severity level in a given dimension, the corresponding dummy is 0. Random effects are estimated at the respondent level.

$$\mu_{ij} = \alpha + \beta t_{ij} + \lambda' x_{ij} t_{ij} + \epsilon_{ij} \qquad (1)$$

A further step is to anchor the coefficients, which allows them to be compared to one another and interpreted meaningfully. This is done, again following Bansback *et al*. (2012), by dividing λ by β for each element of x.

Results

Self-reporting own health

Demographics

In total, 1415 participants took part in the self-report phase. Each variant has just above 200 respondents. Appendix 2 reports the demographic characteristics, demonstrating that the randomisation into versions was successful.

Descriptive statistics on use of composite dimensions

The only logically unambiguous interpretation of the composite dimensions relates to the reporting of "no problems" (See Table 1 for details). In self-reporting their current health, pooling across versions, we find that 47% of participants (n=666) reported no problems with Pain or with Discomfort when these were reported separately, and 51% (n=718) reported no problems with the composite dimension P/D. These proportions are not significantly different according to a chi-square test ($\chi^2(1, N = 2830) = 3.82, p = 0.051$). On the other hand, we do find a significant difference between the composite and separate reporting of no problems for A/D. Specifically, 37% of participants reported no problems with Anxiety or with Depression when separately reported, but 41% reported no problems with the composite A/D ($\chi^2(1, N = 2830) = 4.01, p = 0.045$).

However, with over one third of the sample reporting no problems with anxiety or depression, and almost half reporting no problems with pain or discomfort, current health is clearly not the ideal testbed for examining patterns of self-reporting. Self-reported worst recalled health provides a useful alternative. Pooling across versions, we find that 13% of participants (n=179) reported no problems with Pain or with Discomfort when these were reported separately, whilst 19% reported no problems with the composite dimension P/D. This difference in proportions is strongly statistically significant ($\chi^2(1, N = 2830) = 18.92, p < 0.001$). Similarly, 14% of participants reported no problems with Anxiety or with Depression when separately reported, and 18% reported no problems with the composite A/D ($\chi^2(1, N = 2830) = 6.76, p = 0.01$). Taken together, it appears that even in the unambiguous case of no problems on the composite dimensions, the

composite is not used literally to report problems on one or other component. Appendix 3 reports the full set of cross-tabulations.

Comparing self-reporting rules

Next, we ask what self-reporting rule best fits the observed data, out of the four rules detailed in the study design section. For each composite dimension, we excluded cases where the participant had reported the same value across both components since all models predict the same outcome in these cases. When reporting pain and discomfort, 1111 subjects gave the same ratings when describing their current health, and 901 did so when reporting their worst recalled health. For anxiety and depression, 864 participants gave the same ratings when describing their current health, and 726 when reporting their worst recalled health.

Figure 2 Mean absolute error in predicting the composite dimension from the components, using each of the four self-reporting rules. 95% confidence intervals shown. Higher bars signify worse prediction errors and less support for that hypothesis.

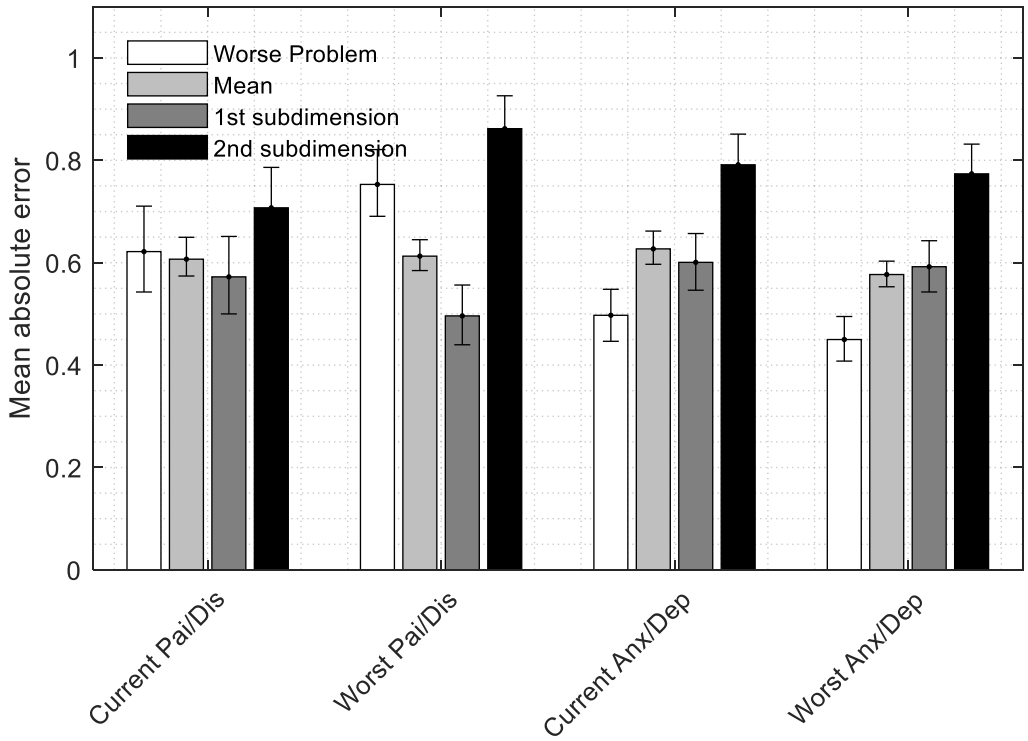


Figure 2 shows the mean absolute prediction error of each self-reporting rule, using the data pooled across the seven versions. In Appendix 4 we report these absolute prediction errors for each version separately. When reporting current P/D, the best predictions come from assuming P/D means Pain (i.e. the first mentioned component rule). This self-reporting rule significantly outperforms both a 'worse component' rule and discomfort alone, though confidence intervals are overlapping with a mean average rule. When worst recalled health is reported, the results strengthen and clearly support the interpretation that P/D is used to report Pain.

For current A/D, the 'worse component' self-reporting rule, in which the composite level is reported at the same as the more severe of the components, is the most accurate in predicting responses. This rule significantly outperforms than the mean rule and the 'depression alone' rule, but confidence intervals overlap with those for the 'anxiety alone' rule. Again, the results strengthen when participants reported their worst recalled health. Here, the 'worse component' rule significantly outperforms all others.

Valuation

A total of 1007 participants completed the valuation phase, and another 18 began the study but failed to complete it. Each variant has 123-149 respondents, and Table A2 in Appendix 2 presents the demographics for each of the seven versions, showing that randomisation resulted in similar demographics across versions. In the practice questions, of the 1007 participants that took part in the study, 47 subjects made a single mistake: 24 in the first choice, 14 in the second and 9 in the third. No participant made more than one mistake.

The data from the follow-up questions on engagement indicate that most respondents did not struggle with completing the study and they tended to find it interesting and clear. The results are presented in Appendix 5.

Discrete choice experiment with duration

The estimated beta coefficients from the regression analyses are presented in Appendix 6.

Across the presentation versions, the coefficients' signs, and the difference in their magnitudes between levels within a given dimension, are as anticipated. Additional years of life significantly increase the likelihood of a scenario being selected across all versions. In almost all cases, problems on a dimension reduce the likelihood that the scenario is selected, compared to having no problems. The exception is Mobility at level 3 (moderate problems) in the Pain only version, where the coefficient has the expected sign but is not significantly different from zero. In all cases, extreme problems (level 5) reduce the chance of selecting the scenario by more than moderate problems (level 3).

Figure 3 Absolute decrements of the anchored coefficients for Pain and/or Discomfort elicited through the DCE_{TTO}, by version. White bars represent moderate problems (level 3) and grey bars represent extreme problems (level 5). 95% confidence intervals shown. Asterisks represent versions where the P/D composite was altered, and so differences might be expected.

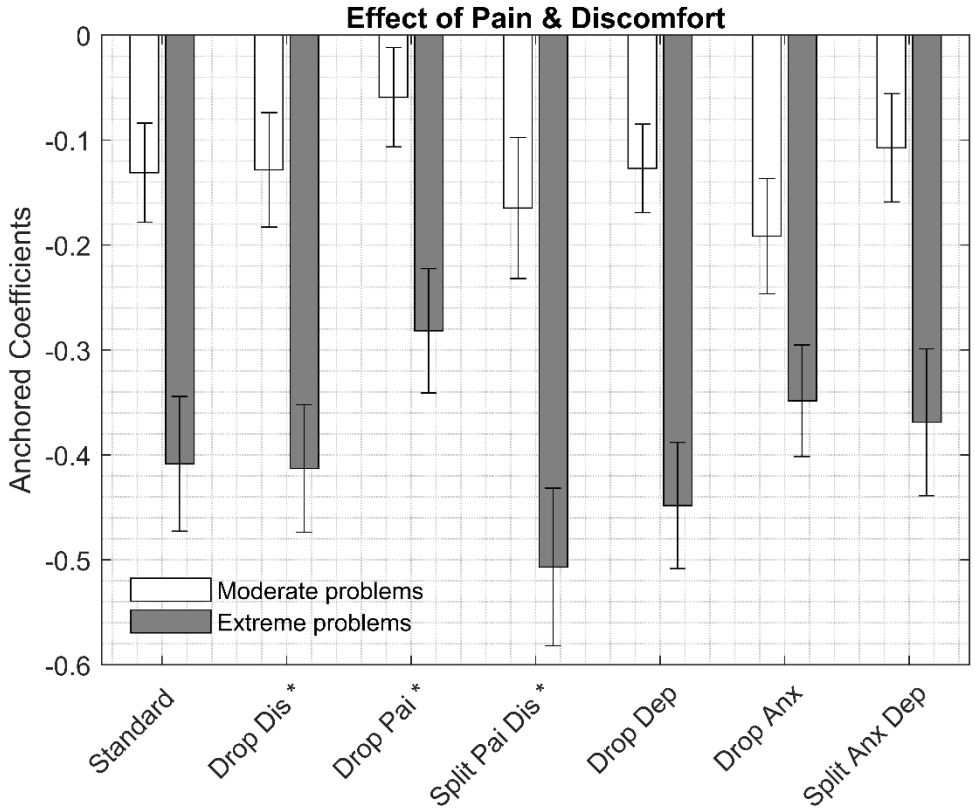
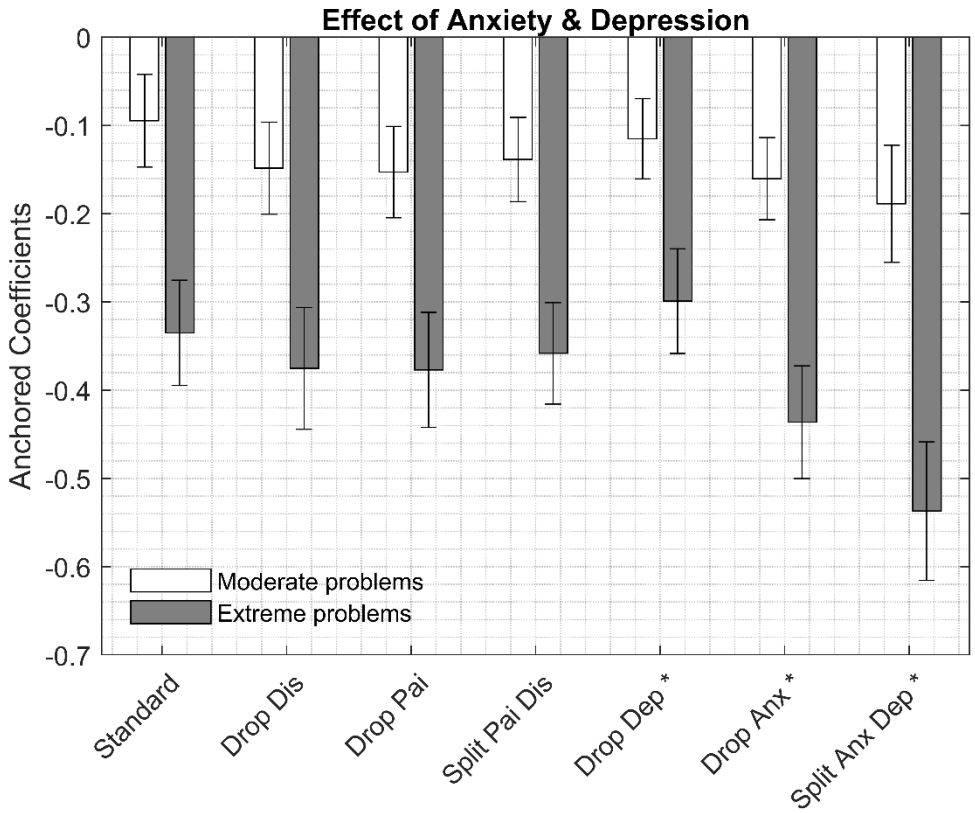


Figure 3 plots the absolute decrements of the anchored coefficients (calculated as λ/β) for the dimensions related to pain and discomfort, and Figure 4 plots the same for the dimensions related to anxiety and depression. Each pair of bars shows anchored coefficients for a version. For example, in the first pair for the standard EQ-5D-5L version, the white bar is the anchored coefficient for the composite P/D at level 3, and the grey bar is the equivalent at level 5. The coefficient for the split treatment (Split Pai Dis) is the sum of the anchored coefficients of the two components, Pain separate and Discomfort separate. Combined standard errors were attained by calculating the combined variance, using: $var(x\&y) = var(x) + var(y) - 2cov(x,y)$.

Figure 4 Absolute decrements of the anchored coefficients for Anxiety and/or Depression elicited through the DCE_{TTO}, by version. White bars represent moderate problems (level 3) and grey bars represent extreme problems (level 5). 95% confidence intervals shown. Asterisks represent versions where the A/D composite was altered, and so differences might be expected.



The heights of the bars can be compared within and across versions. Clearly, level 5 problems are significantly worse than level 3 problems across all versions, and both are significantly worse than the baseline case, no problems. The interesting comparison is between versions. As anticipated, there are no statistically significant differences in the importance of pain or discomfort across the first, fifth, sixth and seventh pairs of bars, since in these versions the P/D composite was presented unaltered. The highest coefficient is that where level 5 Pain and level 5 Discomfort were presented separately, in the Split Pai Dis version.

Comparing the partial drop versions, Drop Dis and Drop Pai, reveals that in valuation, Pain only is clearly perceived to be worse than Discomfort only. Comparing these with the standard presentation reveals that, for levels 5 and 3, the utility decrement from Discomfort only is significantly smaller than the composite P/D, whist the decrements for the composite P/D and for Pain only are indistinguishable. This suggests that when presented with the composite P/D in a valuation exercise, people interpret this as pain.

Turning to Anxiety and Depression in Figure 4, as anticipated we observe no differences between anchored coefficients from the versions where the composite A/D was presented (that is, the first four versions on the diagram). The split treatment where the components were presented separately – i.e. where the coefficient is the sum of the coefficients on the individual components – generates a significantly higher utility decrement compared to the composite, but less than the sum of the two coefficients from the partial drop versions. This holds for both severity levels, 3 and 5. Comparing the partial drop versions and the standard presentation, the level 5 coefficient for Depression only is significantly greater than that for Anxiety only, yet in this case, the composite decrement lies between the decrement for Anxiety and that for Depression.

Discussion

We highlight an inherent logical ambiguity in the EQ-5D system: it is not possible to logically determine how the levels of a composite dimension ought to be interpreted. We clarify the nature of this ambiguity, and provide empirical evidence that explores how the composite dimensions are

actually used in self-report and interpreted in valuation exercises, by systematically altering the presentation of the composite dimensions. The results suggested three key findings:

1) In self-reporting of own health, the use of the composite differs across the dimensions. People appear to use P/D to self-report mainly the level of pain, whilst using A/D to self-report the component out of anxiety and depression for which they have more serious problems. This pattern is most clearly apparent when respondents self-reported their worst recalled health.
2) In valuation of stylised health states, the split versions showed that Pain was clearly perceived to be worse than Discomfort at the same level, and Depression was clearly perceived to be worse than Anxiety at the same level.
3) In valuation, the composite dimension P/D had a utility decrement similar to that for Pain, whilst the decrement for the composite dimension A/D was between that for Anxiety and for Depression.

A recurring insight from our results is that the interpretation of the EQ-5D composite dimensions is sensitive to the dimension of health in question: P/D is differently related to its components Pain and Discomfort than A/D is to Anxiety and to Depression. We find encouraging consistency in how the P/D composite is used and interpreted between task types: in both self-report and in valuation, we found evidence that P/D mainly represents pain. However, we find that A/D is inconsistently used across tasks, with the 'worse problem' interpretation holding for self-report, but not for valuation. This raises questions surrounding whether systematic biases arise when using self-reports in combination with value sets to value conditions involving a mental health detriment.

It is useful to compare these patterns with those found in Tsuchiya et al (2019), which had two null hypotheses relevant for our study. The first was that "the proportion of people who self-report level 1 in a composite dimension is no different from the proportion of people who self-report level 1 in both components when the dimension is decomposed" – in other words, that people use the composite dimensions literally *when self-reporting no problems*. This was not rejected for P/D,

whilst it was for A/D, which is inconsistent with our results. They do not analyse the reporting of having problems at different severity levels. Their second null hypothesis was that in the valuation context "the disutility associated with a composite dimension is no larger than the disutility associated with either component at the same level". For the level 3 coefficients, this hypothesis was not rejected, but for the level 5 coefficients it was rejected. It appears that level 5 P/D is interpreted as extreme pain, while level 5 A/D was interpreted as extreme depression. In contrast, in our case P/D was always interpreted to mean pain, not just at the extreme level. For A/D, the composite lies somewhere in between the two separate component dimensions, which is again inconsistent with the results in Tsuchiya et al. These inconsistencies suggest that the specific patterns may depend on the samples (although both studies used online surveys of the UK public). However, the general conclusion is robust: there is an interacting effect between the context and the composite dimension.

What do our results mean for those using the EQ-5D to measure and value health? Firstly, it suggests there may be a failure to capture some important elements of health states. Specifically, since the P/D dimension in our study appears to be interpreted to mean pain consistently across these two contexts, the composite P/D might fail to capture discomfort either in self-report or in valuation. This is despite the fact that the Partial Drop valuation task resulted in significant reductions in the perceived utility of a health state involving moderate or extreme discomfort. Furthermore, when both pain and discomfort co-occur, this would not be captured by the composite. Nevertheless, if discomfort is generally considered a mild form of pain, this concern is arguably mitigated.

Secondly, it raises difficulties when interpreting the health states that underlie self-reported EQ-5D profiles. The interpretation of A/D is not consistent across tasks. It appears to be interpreted as "the component of Anxiety and Depression with the most severe reported problems" in the self-report context in our study. However, in valuation it appears to be interpreted as "an average of Anxiety and Depression" in our study. This compounds the broader problem, described earlier, that the EQ-

5D is fundamentally incapable of distinguishing between the health of people with different combinations of severity on the sub-dimensions. Our results from the Anx Dep split model clearly suggested that Depression is considered to be more serious (with an anchored disutility of -0.26) than Anxiety at the same level (-0.122), and that having both Anxiety and Depression gives the biggest detriment of all (-0.382). Yet, these differences cannot be inferred in a valuation exercise with the composite dimension. Since a significant minority of our sample in the split versions self-reported problems with Anxiety only, Depression only, or both, this heterogeneity of health states that are indistinguishable from the self-reported composite dimensions is likely to result in serious misallocation of health resources, with a bias towards the undervaluation of Depression, and undervaluation of the co-morbidity of Anxiety with Depression.

The overarching practical implication of the evidence we presented is that care must be taken when identifying health states from people in self-report, and linking these to valuations. Based on our results, the composite dimensions cannot capture the co-occurrence discomfort with pain. Nor can it distinguish between different combinations of severity levels of anxiety and depression. This may result in under-valuation of some health states (for example co-morbidities) and overvaluation of others (for example, self-reported anxiety with no depression being interpreted as depression with no anxiety). More research is required to understand whether splitting both of the composites to create an EQ-7D would be appropriate, especially given the extra burden this would place on valuation studies. An EQ-7D with five levels each will have 78,125 unique health states, a substantial increase from the current 3,125 with EQ-5D-5L (and the 243 with EQ-5D-3L). Valuing such an instrument using a DCE with duration, will involve choice tasks made up of sixteen pieces of information instead of twelve. Furthermore, if (as we contend) discomfort overlaps with the mild end of pain, then the dimensions will not be fully independent and impose restrictions for choice design (for example, no or slight discomfort cannot appear alongside severe or extreme pain), despite the estimated algorithm predicting values for such health states that ought not exist. A realistic proposal for re-organising the EQ-5D must be based on a thorough analysis of the benefits

and costs of doing so (including the additional valuation studies and the management of the transition period in health technology assessment procedures). However, based on the current study, we would suggest that in such a cost benefit analysis EQ-6D, where P/D is split and Discomfort dropped while A/D is split into two separate dimensions, should be given consideration.

We focused on the EQ-5D because of its widespread use, but we believe the concerns raised by our research – and the opportunities for further investigation and improvement of the methodology– apply more broadly to any health (or wellbeing) descriptive system that relies on composite dimensions. These include, but are not limited to, the AQOL-8D (pain or discomfort); 15D (the excretion and mental function dimensions); and SF-6D (the mental health dimension). The issues we raise should also be taken into account when designing new instruments.

Future research could explore the mechanisms underlying the patterns we observe. For instance, work could be done to understand the role of language through studying other language versions of the EQ-5D. Another extension would be to study whether those experiencing mental health problems self-report and value the composite A/D differently. We asked respondents to report their worst remembered health, which goes some way to achieving this, but exploring the same issues with a dedicated sample of this kind would allow us to further generalise our findings.

References

Bansback, N., Brazier, J., Tsuchiya, A., & Anis, A. (2012) Using a discrete choice experiment to estimate health state utility values. *Journal of Health Economics* 31(1), 306-318.

Bansback, N., Hole, A. R., Mulhern, B., & Tsuchiya, A. (2014) Testing a discrete choice experiment including duration to value health states for large descriptive systems: Addressing design and sampling issues. *Social Science and Medicine* 114, 38-48

Brazier J, Rowen D, Tsuchiya A, Yang Y, & Young TA. (2011) The impact of adding an extra dimension to a preference-based measure. *Social Science and Medicine*, 73, 245–53.

Brooks R. (1996) EuroQol: The current state of play. *Health Policy*. 37, 53–72.

Bryan, S., Jowett,S., Longworth, L., & Pickard, S. (2005). Does the EQ-5D "anxiety/depression" item measure anxiety, depression, both or neither? *HESG working paper*

Devlin N, Shah K, Mulhern B, Pantiri K, & van Hout B (2017), A New Valuation Method: Directly Eliciting Personal Utility Functions, *Office of Health Economics Research Paper* 17/06

Golicki, D., & Niewada, M. (2017). EQ-5D-5L Polish population norms. Archives of medical science: AMS, 13(1), 191.

Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, & Badia X (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*. 20(10),1727–36.

Hinz A, Kohlmann T, Stöbel-Richter Y, Zenger M & Brähler E. (2014) The quality of life questionnaire EQ-5D-5L: psychometric properties and normative values for the general German population. *Quality of Life Research*. 23, 443–447.

Krabbe, P. F. M., Stouthard, M. E. A., Essink-Bot, M. & Bonsel, G. J. (1999) The effect of adding a cognitive dimension to the EuroQol multiattribute health-status classification system. *Journal of Clinical Epidemiology* 52(4), 293-301

Longworth L, Yang Y, Young T, Mulhern B, Hernández Alava M, Mukuria C, Rowen D, Tosh J, Tsuchiya A, Evans P, Devianee Keetharuth A, & Brazier J. (2014) Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technology Assessment*. 18(9),1-224.

McCaffrey N, Kaambwa B, Currow DC & Ratcliffe J. (2016) Health-related quality of life measured using the EQ-5D-5L: South Australian population norms. *Health and Quality of Life Outcomes*. 14, 133.

McDonald, R & Mullett, T. L., Experimental Tests of the Robustness of Health Decisions Using the EQ-5D (April 4, 2020). Available at SSRN: https://ssrn.com/abstract=3568488

Macran S, & Kind P. (2000) EQ-5D Valuations from a British national Postal survey. In: Cabase´s JM, Gaminde I (eds), *17th Plenary Meeting of the EuroQol Group Discussion Papers*, Pamplona, Spain. https://eq-5dpublications.euroqol.org/download?id=0_53518&fileId=53941

Melzack, R. (1975). The McGill Pain Questionnaire: major properties and scoring methods. *Pain*, 1(3), 277-299.

Melzack, R. (1987). The short-form McGill pain questionnaire. *Pain*, 30(2), 191-197.

Mulhern, B., Norman, R., Street, D.J. & Viney, R., (2019). One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. *PharmacoEconomics*, 37(1), 29-43.

Mulhern, B., Norman, R. & Shah, K. (2018) How should DCE with duration choice sets be presented for the valuation of health states? *Medical Decision Making* 38(3), 306-318

Mulhern, B. Norman, R., Lorgelly, P., Lancsar, E., Ratcliffe, J., Brazier, J., & Viney, R. (2017) Is dimension order important when valuing health states using discrete choice experiments including duration? *Pharmacoeconomics* 35(4), 439-451

Mulhern, B. Bansback, N., Brazier, J., Buckingham, K., Cairns, J., Devlin, N., Dolan, P., Hole, A. R., Kavetsos, G., Longworth, L. Rowen, D., & Tsuchiya, A. (2014) Preparatory study for the revaluation of the EQ-5D tariff: methodology report. *Health Technology Assessment*. 18(12), 1-191

Ngene 1.1.2 User Manual & Reference Guide, 2014 http://www.choice-metrics.com/documentation.html (accessed 5th May 2017)

Tsuchiya, A., Bansback, N., Hole, A. R., & Mulhern, B. (2019). Manipulating the 5 Dimensions of the EuroQol Instrument: The Effects on Self-Reporting Actual Health and Valuing Hypothetical Health States. *Medical Decision Making*, 39(4), 380-392.

Wolfs, C. A. G., Dirksen, C. D., Kessels, A., Willems, D. C. M., Verhey, F. R. J. & Severens, J. L. (2007) Performance of the EQ-5D and the EQ-5D+C in elderly patients with cognitive impairments *Health and Quality of Life Outcomes*. 5(33)

Yang, Y., Brazier, J., Tsuchiya, A. (2013) Effect of adding a sleep dimension to the EQ-5D descriptive system: A "bolt-on" experiment. *Medical Decision Making* 34(1), 43-53

Acknowledgements

This work was supported by the Economic and Social Research Council [grant numbers ES/K002201/1, ES/P008976/1] Network for Integrated Behavioural Science. The authors would like to thank the members of that network for their critique of earlier drafts. Thanks also to Arne Risa Hole for advice on the DCE design, the Health Economics Research Unit at the University of Aberdeen for valuable feedback, and to participants and discussants at the Health Economics Study Group meeting in Bristol (June 2018) and the EuroQol Group meeting in Lisbon (Sep 2018)

Appendices

Appendix 1: instructions for discrete choice experiment questions, standard presentation

Figure A1

You will be presented with two imaginary health "scenarios". Each scenario describes a different health state and we want you to imagine what it would be like to live in each one. There is also information on how long you will live in each health state. After that time, death will be swift and painless.

You will be asked which health scenario you think is better.

As you have already seen, five areas of health will be included in each overall health state. The areas of health are:

- Mobility (walking about)
- Self care (washing or dressing yourself)
- Usual activities (e.g. work, study, housework, family or leisure activities)
- Pain / Discomfort
- Anxiety / Depression

Please imagine you will experience each health state for the period shown without further relief or treatment.

Please also imagine that you will have no other health problems besides what is indicated.

There are no right or wrong answers to any of the questions. We are interested in what you think.

The next pages have 4 practice questions. Please complete these in your own time.

Appendix 2: demographics and randomisation into versions

Demographics for self-report sample

Table A1 Demographics and randomisation into versions, self-report

Demographic	Standard	Drop Dis	Drop Pai	Split Pai Dis	Drop Dep	Drop Anx	Split Anx Dep	Pooled
Female (%)	58.2%	51.7%	61.3%	55.9%	55.4%	60.1%	61.4%	57.7%
Age in years (Mean (S.D.))	37.1 (11.9)	36.6 (12.9)	37.9 (13.7)	36.0 (13.0)	38.0 (12.5)	37.4 (12.3)	40.2 (13.0)	37.6 (12.8)
Employed full or part time (%)	70.1%	59.6%	60.4%	63.4%	66.8%	66.5%	65.8%	64.7%
Educated beyond school leaving age (%)	90.0%	88.7%	86.6%	92.1%	85.6%	86.7%	87.1%	88.1%
Degree or equivalent (%)	69.2%	58.1%	58.4%	62.4%	55.9%	59.1%	60.9%	60.6%
Responsible for children under 18y (%)	35.8%	26.1%	31.2%	25.2%	32.7%	34.5%	38.1%	31.9%
Had serious illness (%)	30.3%	26.6%	30.7%	27.7%	33.7%	33.0%	34.7%	31.0%
Self reported health:	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Excellent (%)	7.5%	10.8%	7.9%	12.9%	10.4%	10.3%	13.4%	10.5%
Very good (%)	37.3%	33.0%	32.2%	36.1%	34.2%	32.5%	37.1%	34.6%
Good (%)	29.9%	32.0%	36.6%	29.2%	31.7%	33.5%	29.7%	31.8%
Fair (%)	16.4%	17.7%	17.3%	14.4%	16.3%	17.7%	14.9%	16.4%
Poor (%)	9.0%	6.4%	5.9%	7.4%	7.4%	5.9%	5.0%	6.7%
Total n per treatment	201	203	202	202	202	203	202	1415

Demographics for valuation sample

Table A2 Demographics and randomisation into versions, valuation

Demographic	**Standard**	**Drop Dis**	**Drop Pai**	**Split Pai Dis**	**Drop Dep**	**Drop Anx**	**Split Anx Dep**	**Pooled**
Female (%)[1]	58.7%	58.0%	42.3%	56.1%	50.4%	56.7%	58.4%	54.6%
Age in years (Mean (S.D.))	32.1 (7.3)	32.1 (10.4)	29.6 (8.1)	30.5 (8.8)	30.5 (8.9)	31.3 (7.8)	31 (7.6)	31 (8.5)
Employed full or part time (%)	60.1%	57.2%	52.0%	60.1%	59.4%	63.3%	65.8%	59.9%
Educated beyond school leaving age (%)	85.5%	88.4%	93.5%	93.9%	82.0%	89.2%	89.3%	88.8%
Degree or equivalent (%)	65.2%	70.3%	65.0%	67.6%	59.4%	63.3%	62.4%	64.8%
Responsible for children under 18y (%)	54.3%	41.3%	31.7%	33.1%	30.1%	38.3%	42.3%	38.9%
Had serious illness (%)	33.3%	33.3%	30.1%	31.1%	27.8%	24.2%	28.2%	29.8%
Self reported health:								
Excellent (%)	12.3%	11.6%	8.1%	8.8%	10.5%	8.3%	14.1%	10.6%
Very good (%)	37.0%	37.7%	43.1%	48.0%	49.6%	35.8%	43.6%	42.3%
Good (%)	31.2%	35.5%	39.8%	27.0%	27.1%	30.8%	28.2%	31.2%
Fair (%)	15.9%	12.3%	6.5%	12.8%	10.5%	16.7%	11.4%	12.3%
Poor (%)	3.6%	2.9%	2.4%	3.4%	2.3%	8.3%	2.7%	3.6%
Total n per treatment	138	138	123	148	133	120	149	949

[1] A total of 3 participants gave non-binary gender descriptions. One each in the conditions Standard, Drop Pai, and Drop Dep.

Appendix 3: Self-report cross-tabulations and supplementary analyses

Table A3 Use of composite and separate dimensions to report current pain and discomfort.

	None	Slight	Moderate	Severe	Extreme
P/D composite	718 (51%)	518 (37%)	132 (9%)	36 (3%)	11 (1%)

Pai separate	Dis separate None	Slight	Moderate	Severe	Extreme
None	666 (47%)	123 (9%)	10 (1%)	6 (0%)	1 (0%)
Slight	73 (5%)	338 (24%)	27 (2%)	3 (0%)	0 (0%)
Moderate	3 (0%)	24 (2%)	83 (6%)	12 (1%)	2 (0%)
Severe	5 (0%)	3 (0%)	8 (1%)	18 (1%)	1 (0%)
Extreme	0 (0%)	1 (0%)	1 (0%)	1 (0%)	6 (0%)

Table A4 Use of composite and separate dimensions to report current anxiety and depression

	None	Slight	Moderate	Severe	Extreme
A/D composite	581 (41%)	494 (35%)	237 (17%)	75 (5%)	28 (2%)

Anx separate	Dep separate None	Slight	Moderate	Severe	Extreme
None	529 (37%)	54 (4%)	12 (1%)	1 (0%)	0 (0%)
Slight	231 (16%)	232 (16%)	36 (3%)	2 (0%)	0 (0%)
Moderate	48 (3%)	77 (5%)	77 (5%)	13 (1%)	3 (0%)
Severe	14 (1%)	13 (1%)	13 (2%)	18 (1%)	5 (0%)
Extreme	1 (0%)	3 (0%)	3 (0%)	9 (1%)	8 (1%)

Table A5 Use of composite and separate dimensions to report worst recalled pain and discomfort.

	None	Slight	Moderate	Severe	Extreme
P/D composite	263 (19%)	306 (22%)	435 (31%)	288 (20%)	123 (9%)

Pai separate	Dis separate None	Slight	Moderate	Severe	Extreme
None	179 (13%)	76 (5%)	44 (3%)	13 (1%)	3 (0%)
Slight	25 (2%)	170 (12%)	93 (7%)	24 (2%)	5 (0%)
Moderate	5 (0%)	26 (2%)	268 (19%)	97 (7%)	17 (1%)
Severe	1 (0%)	3 (0%)	31 (2%)	192 (14%)	42 (3%)
Extreme	0 (0%)	0 (0%)	0 (0%)	9 (1%)	92 (7%)

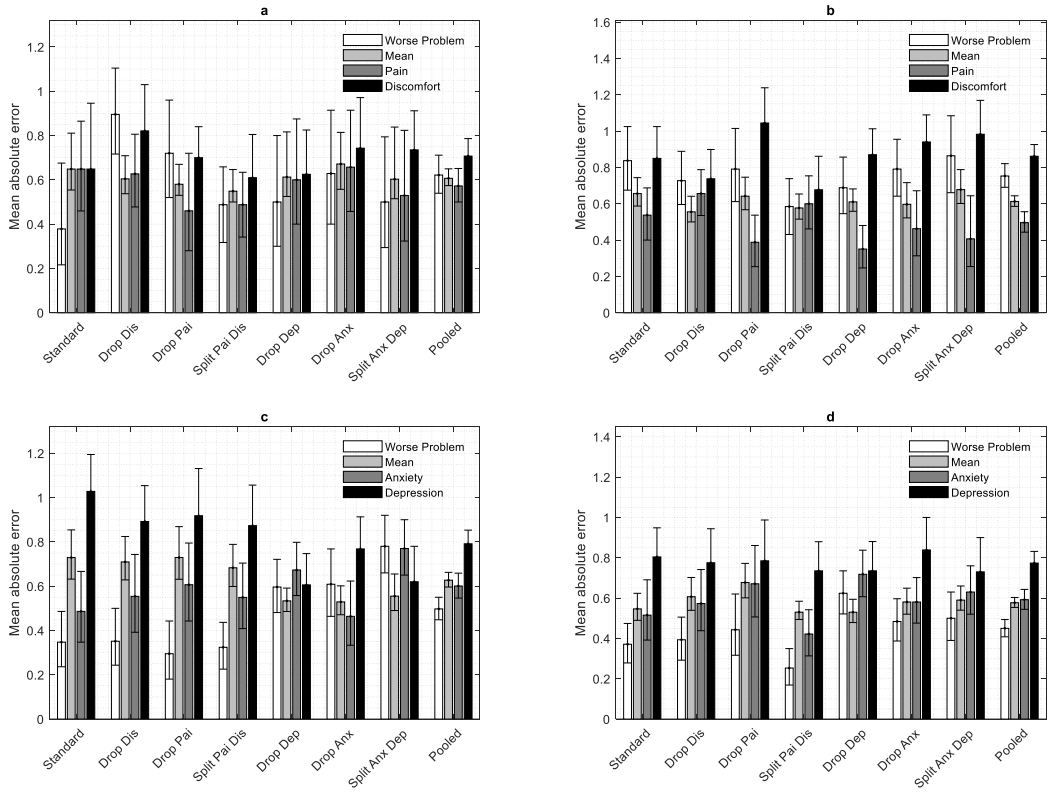
Table A6 Use of composite and separate dimensions to report worst recalled anxiety and depression.

	None	Slight	Moderate	Severe	Extreme
A/D composite	255 (18%)	349 (25%)	365 (26%)	244 (17%)	202 (14%)

Anx separate	Dep separate None	Slight	Moderate	Severe	Extreme
None	204 (14%)	29 (2%)	10 (1%)	2 (0%)	1 (0%)
Slight	121 (9%)	142 (10%)	38 (3%)	3 (0%)	2 (0%)
Moderate	48 (3%)	127 (9%)	174 (12%)	29 (2%)	9 (1%)
Severe	19 (1%)	36 (3%)	85 (6%)	95 (7%)	24 (2%)
Extreme	4 (0%)	17 (1%)	25 (2%)	60 (4%)	111 (8%)

Appendix 4: Full results of comparison of reporting rules

Figure A2



Appendix 5: Results of engagement questionnaire

Table A7 Results relating to respondents' impressions of the study, valuation phase

Statement	Proportion selecting this statement
Too many tasks	10.9%
I could answer 5 or 6 more of the choices	18.4%
I got tired half way through	12.1%
Difficult to distinguish between the scenarios	20.1%
Some of the health states seemed very unlikely	26.8%
Interesting survey	62.2%
Boring	6.6%
The task being asked is clear	60.3%
Not sure about my answers	9.2%
Confident about my answers	47.4%

Appendix 6: Regression results from valuation task

Table A8 Unanchored estimated beta coefficients from logistic regression of the DCE_{TTO} data

Probability option chosen	Standard	Drop Dis	Drop Pai	Split Pai Dis	Drop Dep	Drop Anx	Split Anx Dep
Intercept	-0.386	-0.423	-0.384	0.172	0.121	-0.498	0.121
Mo = 3 x dur	-0.156**	-0.085	-0.295***	-0.166***	-0.176***	-0.167**	-0.177***
Mo = 5 x dur	-0.541***	-0.473***	-0.668***	-0.5***	-0.425***	-0.528***	-0.498***
SC = 3 x dur	-0.13**	-0.132*	-0.143**	-0.06	-0.113*	-0.191***	-0.193***
SC = 5 x dur	-0.443***	-0.586***	-0.486***	-0.417***	-0.408***	-0.54***	-0.364***
UA = 3 x dur	-0.192***	-0.182***	-0.215***	-0.122*	-0.147**	-0.302***	-0.103*
UA = 5 x dur	-0.441***	-0.512***	-0.535***	-0.463***	-0.425***	-0.458***	-0.387***
P/D = 3 x dur	-0.267***				-0.27***	-0.422***	-0.217***
P/D = 5 x dur	-0.832***				-0.953***	-0.767***	-0.746***
Pai = 3 x dur		-0.269***		-0.225***			
Pai = 5 x dur		-0.865***		-0.703***			
Dis = 3 x dur			-0.129*	-0.1*			
Dis = 5 x dur			-0.616***	-0.296***			

A/D = 3 x dur	-0.193***	-0.311***	-0.334***	-0.273***			
A/D = 5 x dur	-0.682***	-0.786***	-0.824***	-0.706***			
Anx = 3 x dur					-0.245***		-0.122**
Anx = 5 x dur					-0.636***		-0.383***
Dep = 3 x dur						-0.353***	-0.26***
Dep = 5 x dur						-0.961***	-0.704***
Duration	2.036***	2.093***	2.187***	1.971***	2.126***	2.202***	2.023***
No. observations	8280	8280	7380	8880	7980	7200	8940
Log Likelihood	1785	2422	2254	2578	1838	1840	3373
BIC	-3447	-4721	-4388	-5011	-3554	-3561	-6602